



TITLE:

# Line monitoring by near-infrared chemometric technique for potential ethanol production from hydrothermally treated *Eucalyptus globulus*

AUTHOR(S):

Horikawa, Yoshiki; Imai, Makiko; Kanai, Keiko; Imai, Tomoya; Watanabe, Takashi; Takabe, Keiji; Kobayashi, Yoshinori; Sugiyama, Junji

CITATION:

Horikawa, Yoshiki ...[et al]. Line monitoring by near-infrared chemometric technique for potential ethanol production from hydrothermally treated *Eucalyptus globulus*. *Biochemical Engineering Journal* 2015, 97: 65-72

ISSUE DATE:

2015-05

URL:

<http://hdl.handle.net/2433/196079>

RIGHT:

© 2015 Elsevier B.V. NOTICE: this is the author's version of a work that was accepted for publication in *Biochemical Engineering Journal*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Biochemical Engineering Journal*, 97 (2015), doi:10.1016/j.bej.2015.01.003; This is not the published version. Please cite only the published version.; この論文は出版社版ではありません。引用の際には出版社版をご確認ください。

## Line monitoring by near-infrared chemometric technique for potential ethanol production from hydrothermally treated *Eucalyptus globulus*

Yoshiki Horikawa,<sup>1\*</sup> Makiko Imai,<sup>1</sup> Keiko Kanai,<sup>1</sup> Tomoya Imai,<sup>1</sup> Takashi Watanabe,<sup>1</sup> Keiji Takabe,<sup>2</sup> Yoshinori Kobayashi,<sup>3</sup> Junji Sugiyama<sup>1</sup>

<sup>1</sup>Research Institute for Sustainable Humanosphere, Kyoto University, Kyoto, Japan

<sup>2</sup>Division of Forest and Biomaterials Science, Graduate School of Agriculture, Kyoto University, Kyoto, Japan

<sup>3</sup>Tsukuba Research Laboratory, Japan Bioindustry Association, Ibaragi, Japan

Correspondence to: Yoshiki Horikawa, Research Institute for Sustainable Humanosphere, Kyoto University, Uji, Kyoto 611-0011, Japan.

E-mail: yhorikawa@rish.kyoto-u.ac.jp

Tel: +81-774-38-3634; Fax: +81-774-38-3635

## Abstract<sup>1</sup>

This study reports a method that combines near-infrared (NIR) measurements with multivariate analysis to predict the saccharification efficiency of hydrothermally pretreated *Eucalyptus globulus* during ethanol conversion. Optimization of the NIR data with or without spectral treatment determined the best calibration model in the region 10000–4000  $\text{cm}^{-1}$  of the original spectra, with an RMSEP of 2.08% and  $R_p^2$  of 0.99. By investigating the regression coefficient to understand the key regions and chemical components, for original and multiplicative scatter correction (MSC)-treated spectra, the water absorption and higher wavenumber regions were important. For the second derivative spectra, the regression model was constructed based on the CH overtone vibrations (6050–5500  $\text{cm}^{-1}$ ). The regression coefficient demonstrated that the removal of hemicellulose resulted in higher lignin content, which might affect the biomass properties in terms of water absorption and enhanced enzymatic hydrolysis evaluated by dinitrosalicylic acid (DNS) method. For a higher throughput system, aqueous sample analysis was performed using an immersion probe equipped with an InGaAs detector, which generated an acceptable calibration model having RMSEP of 4.25 % and  $R_p^2$  of 0.94. These results show the great potential of NIR spectroscopy for achieving fast, accurate, and nondestructive analysis, and its highly adaptability for maintaining an ethanol bioconversion system.

**Keywords:** NIR spectroscopy; Pretreatment; Cellulase; Cellulose; Lignin; Bioprocess monitoring

---

<sup>1</sup> near-infrared (NIR), multiplicative scatter correction (MSC), filter paper activity (FPA), dinitrosalicylic acid (DNS), triglycine sulfate (TGS), partial least squares (PLS), root-mean-square error of prediction (RMSEP), ratio of performance to deviation (RPD), root-mean-square error of calibration (RMSEC)

## 1. Introduction

The chemical composition and ultrastructure of plant cell walls greatly influence their enzymatic digestibility during ethanol conversion, a process which is of increasing interest due to energy shortages from the excessive consumption of non-renewable fuels. A line-monitoring system, defined as a rapid and reproducible analytical method, is needed to support a stable conversion process; current monitoring uses conventional wet chemical analyses that involve many steps and employ additional reagents, increasing the environmental load, expense, and time requirements. Near-infrared spectroscopy is a fast, non-destructive, and powerful method for the characterization of organic materials. It provides information about the structure, chemical microenvironment of the constituent biomolecules, and their functional group distributions. Its combination with multivariate analysis, known as chemometrics, enables the simultaneous quantification of chemical components as well as assessment of the physical properties of plant materials [1–2].

This quantitative method has been applied to biomass samples, as initially reported by Sanderson et al. [3], who demonstrated the establishment of a calibration model between NIR spectra acquired in the reflectance mode and chemical composition, such as individual carbohydrate, lignin, and ash contents. Later studies investigated calibration models for systems employing various NIR spectral pretreatments [4–5] alone or in combination with other multivariate analysis methods such as PCA [6] and SIMCA [7], or the interpretation of regression vectors or loading factors [8]. For chemically pretreated rice straw and *Erianthus*, we previously demonstrated a chemometric approach to the saccharification ratio, which is an important index for assessing pretreatment methods [9–10]. All of the NIR spectra conducive to the acceptable calibration model were measured from residues that had strongly NIR-light absorbing and scattering properties, with a few sample preparation steps such as drying or molding. Therefore, this NIR measurement was limited to “at-line monitoring” in terms of a manufacturing control system for ethanol conversion. On the other hand, liquid sample measurement would enable direct screening without dilution or extensive sample preparation, which promises higher throughput analysis corresponding to “in-line monitoring.” Needless to say, the use of NIR spectroscopy for aqueous analysis has been difficult because the signals of interest are obscured by strong and broad vibrational bands due to water [11]. However, this challenge has been overcome by tweaking measurement conditions or spectral instrumentation. In a successful example, Rodriguez-Saona et al. [12] constructed a calibration model using different sampling devices, ultimately finding that the transmittance mode achieved the rapid analysis of polysaccharides in fruit juices by NIR spectroscopy.

Cozzolino et al. [13] also created an acceptable calibration model for the phenolic compounds in red wine fermentation by NIR spectroscopy in the transmittance mode. Liebmann and co-workers [14–15] applied a fiber optic probe to successfully predict components such as glucose and ethanol in liquid samples during fermentation.

The present study initially reports the development of a high-throughput analysis using NIR spectroscopy to predict the digestibility of hydrothermally treated *Eucalyptus globulus* from hydrolyzed residues. The spectral pretreatment and region were optimized to construct a better regression model. Next, to more fully understand the relationships between the NIR spectra and enzymatic hydrolysis toward efficient ethanol conversion, a regression coefficient was calculated and interpreted in terms of chemical and structural components. Finally, for a higher throughput system for practical bioprocess monitoring, we tried to construct a liquid analysis method by using an immersion probe which could be directly inserted in a filtrate after pretreatment.

## 2. Materials and Methods

### 2.1. Sample preparation

*E. globulus* wood chips, kindly provided by the Oji Paper Co., Ltd. (Tokyo Japan), were subjected to two-stage milling. Stage one involved rough milling using an Orient mill VM-16 (Seishin Enterprise Corp., Tokyo, Japan), followed by a second milling with a Bantam mill AP-BL (Hosokawa Micron Corp., Osaka, Japan) to achieve particles that would pass through a 150- $\mu$ m screen. The products were pretreated hydrothermally with a Biotage® Initiator™ Sixty (Japan) under different conditions of temperature (160–200°C) and reaction time (1–64 min) in order to widely vary the saccharification ratio. The solid samples were repeatedly washed with distilled water for the subsequent enzymatic hydrolysis and NIR spectroscopic analysis. The filtrate was also collected for NIR spectral analysis.

### 2.2. Enzymatic hydrolysis

A commercial enzyme cocktail, Accellerase 1500 (Genencor, Danisco US, Inc., Rochester, NY), was used for saccharification throughout this study. Enzymatic saccharification was performed with the enzyme (40 FPU) in 100 mM acetate buffer (4 mL, pH 5.0) with a 5% (w/v) dry pretreated substrate loading at 50°C with 150 strokes/min for 24 h. The filter paper activity (FPA) was measured according to the standard procedure recommended by the NREL [16]. The amount of sugar released was measured as a reducing sugar by employing the dinitrosalicylic acid (DNS) method [17]. The saccharification ratio (%) was estimated by the

following equation:

$$\text{Saccharification ratio (\%)} = [\{\text{Reducing sugar (mg/mL)} \times 4 \text{ (mL)} \times 0.9\} / 200 \text{ (mg)}] \times 100$$

### 2.3. NIR data acquisition

For a solid sample after hydrothermal pretreatment, a disk was cast using a hand press after collecting around 0.04 g sample, according to a published method [9]. NIR spectra were recorded on a PerkinElmer Spectrum 100N system equipped with an integrating sphere diffuse reflectance accessory (PerkinElmer) and a triglycine sulfate (TGS) detector. Two spectra were obtained from each sample placed directly onto this accessory and recorded at a resolution of  $16 \text{ cm}^{-1}$  with an acquisition of 32 scans. Liquid samples were analyzed with an immersion probe (Hellma Analytics) having a path length of 1 mm with an indium-gallium-arsenide (InGaAs) detector. NIR spectra were recorded in triplicate when measuring using an immersion probe, which was inserted directly into the filtrate at  $25^\circ\text{C}$  for a few seconds, with a spectral resolution of  $16 \text{ cm}^{-1}$  and a scan acquisition number of 64. After each measurement, the immersion probe was washed several times with distilled water and cleaned using a blower. The original spectra were subjected to multiplicative scatter correction (MSC) or secondary differentiation using the 9-point Savitzky–Golay algorithm [18] before multivariate analysis.

### 2.4. Chemometric analysis

A total of 72 samples was divided into sets of 54 for calibration and 18 for prediction (Table 1). The partial least squares (PLS) regression applied to construct the calibration model for the saccharification ratio was employed using commercial software (Unscrambler v. 9.8; CAMO Software, Inc., Woodbridge, NJ). The number of PLS factors was optimized by full cross-validation, where a single sample was kept out of the model and then predicted by the constructed model without the sample. This process was repeated for every individual sample in the calibration set. The coefficient of determination for calibration ( $R_c^2$ ) and root-mean-square error of calibration (RMSEC) were used to evaluate calibration performance. The models developed were assessed by the coefficient of determination for prediction ( $R_p^2$ ), the root-mean-square error of prediction (RMSEP), and the ratio of performance to deviation (RPD). A larger value of RPD reflects the greater accuracy of the data fitted by the calibration; furthermore, an RPD that exceeds 2.5 is considered satisfactory for screening [19].

### 3 Results and discussion

#### 3.1. Construction of PLS calibration model from solid samples

Hydrothermal pretreatments were performed using the Biotage® Initiator™ Sixty system, which differs from an autoclave-type in that it can reach a preset temperature quickly because of its microwave heating system. The statistical summary of the saccharification ratio is given in Table 1, in which a wide variance is shown in both the calibration and prediction sets for randomly selected samples.

The solid sample was initially examined for the construction of a calibration model. The disk was made from the residue after pretreatment because, for stable and repeatable data measurements via an integrating sphere diffuse reflectance accessory equipped with TGS detector, it is necessary to employ uniform particle sizes. The original spectra obtained (Fig. 1a) were transformed by MSC (Fig. 1b) or by second differential processing (Fig. 1c), both of which are often employed for better regression modeling.

To determine the optimal spectral region for constructing the PLS regression model with higher performance, the full-length NIR range of 10000–4000  $\text{cm}^{-1}$ , along with the different electromagnetic wave absorptions of the functional groups, was separated into four ranges based on the properties of the molecular vibrations: (i) 10000–7300  $\text{cm}^{-1}$ , in which the second and third overtones are present; (ii) 7300–6050  $\text{cm}^{-1}$ , in which mainly the OH overtone vibrations are detected; (iii) 6050–5500  $\text{cm}^{-1}$ , where CH vibrations and aromatic framework vibrations are detected; and (iv) 5500–4000  $\text{cm}^{-1}$ , which corresponds to several combinational vibrations. These respective regions and their combinations generated the PLS calibration models for the saccharification ratios summarized in Figure 2. At first, we investigated the construction of the regression model from the original spectra. Optimization of the NIR regions determined that the full length gave the best calibration performance with an RMSEP of 2.08%,  $R_p^2$  of 0.99, and RPD of 8.55 (Figure 2a). When regression models

were obtained from each single NIR region, 10000–7300  $\text{cm}^{-1}$  showed good performance. Next, in the case of the MSC process, the PLS factor numbers were decreased in comparison with the original spectra (Table 2b), and the best model was also created based on the full length (Figure 2b). Interestingly, improvement of the regression model through the MSC process depended on the spectral region; the OH overtone vibration in the range of 7300–6050  $\text{cm}^{-1}$  clearly increased the regression performance. On the other hand, the regions of 10000–7300  $\text{cm}^{-1}$  and 5500–4000  $\text{cm}^{-1}$  were nearly unaltered. Third, for the second derivative spectra, the regression performance became lower in the range of 10000–7300  $\text{cm}^{-1}$ , wherein robust calibration models were obtained from the original and MSC-treated spectra (Table 2c). As shown in Figure 1c, the important signals in the 10000–7300  $\text{cm}^{-1}$  region disappeared by the second derivative process, which might have decreased the predictive performance. Furthermore, different from the original and MSC-treated spectra, the best calibration model was given from the comparatively narrow region of 6050–5500  $\text{cm}^{-1}$ , wherein  $R_p^2$  and RMSEP showed values of 0.98 and 2.23 (Figure 2c). In comparison with our previous papers [9–10], the obtained calibration model demonstrated quite reliable predictions, though this study focused on only hydrothermal treatment.

### 3.2. Interpretation of calibration model from solid analysis for hydrothermal pretreatment

In order to investigate what is directly or indirectly a key factor for enzymatic hydrolysis and why the important NIR regions depend on the spectral process, we calculated the regression coefficient presented in Figure 3. For interpretation of the regression coefficient, the NIR bands were referenced to recent reports [20–23]. The patterns of the original spectra were similar to those of the MSC-treated spectra, in that the negative bands at 7108 and 5224  $\text{cm}^{-1}$  were intensified (Figure 3a and b). These bands were reported to be specific to water absorption, which indicated that the pretreatment with the lower amount of



absorbed water resulted in higher digestibility. In addition, a stronger absorbance was seen at a higher wavenumber, approaching  $10000\text{ cm}^{-1}$ . In NIR spectroscopy, it is known that a dark sample absorbs higher wavenumber vibrations [24]. Therefore, this demonstrated that the dark colored sample became more highly digestible by cellulase. In the case of the second derivative spectra, small bands were seen in the region slightly above  $7000\text{ cm}^{-1}$ , which were consistent with the lower performance model obtained from the  $10000\text{--}7300\text{ cm}^{-1}$  region (Table 2c). In the range of  $6050\text{--}5500\text{ cm}^{-1}$ , where the best calibration model was obtained, a noticeable positive band, characteristic of hemicellulose, was seen at  $5810\text{ cm}^{-1}$ . Moreover, a negative band at  $5980\text{ cm}^{-1}$ , attributable to aromatic framework vibrations, was detected, and the band relating to lignin at  $6900\text{ cm}^{-1}$  also was negative. These results indicated that pretreatments that resulted in lower hemicellulose and higher lignin contents increased the saccharification ratio, which is consistent with a report published by our co-author Kobayashi [25]. These discussions on the regression coefficient can be summarized as follows: the hydrothermal treatment preferentially removes hemicellulose and results in higher lignin contents, as already reported [26–27], which leads to hydrophobic properties. Thus, the sample after this treatment tends to contain less absorbed water. In addition, the comparatively higher lignin contents results in a darker sample color derived by strongly hydrothermal treatment. Accordingly, the removal of hemicellulose is directly correlated with the saccharification ratio by increasing enzymatic accessibility to cellulose, whereas the lignin content, water absorption, and color change are indirect influences for the construction of a calibration model to predict digestibility.

### 3.3. Construction of PLS calibration model from liquid samples

As mentioned, NIR spectroscopy can be applied to predict digestibility in a pretreated residue. However, in this protocol, sample preparation such as drying and tablet formation is needed. For a more rapid, simple, and practical line monitoring system, we

examined the possibility of liquid spectral analysis. The sphere diffuse reflectance accessory coupled with the TGS detector was initially tested, using a glass cell for the filtrates recovered during the successful analysis of solid samples. However, all the calibration models resulted in poor performance even though various liquid quantities, NIR regions, and spectral pretreatments were evaluated (data not shown). Therefore, we applied a transmittance-type immersion probe equipped with an InGaAs detector suitable for liquid analysis because of its high sensitivity and low noise. Similarly to the solid analysis study, the original spectra and two pretreated spectra were investigated to create a better calibration model. As shown in Figures 1c–e, large vibrations relating to water were seen near 7000 and 5000  $\text{cm}^{-1}$ ; in the latter region, absorbance saturation occurred. Therefore, the NIR region 5400–4900  $\text{cm}^{-1}$  was omitted throughout the liquid analysis. Table 3 summarizes the calibration models based on the original, MSC, and second derivative spectra from the NIR regions used during the liquid sample analysis. For the original and MSC-treated spectra, the full length generated best regression model (Figure 4a and b), in common with the results of the solid analysis. The regression model from the original spectra over 10000–4000  $\text{cm}^{-1}$  provided values of  $R_p^2$ , RMSEP, and RPD of 0.94, 4.45, and 3.99, demonstrating great potential for the incorporation of the method into a manufacturing system as an in-line monitoring technique for better ethanol production. Liebmann et al. [15] successfully reported calibration models using a fiber optic probe to predict glucose, ethanol, glycerol, and other components, in which the values of  $R_p^2$  were more than 0.94. The models developed in this study are comparable to the published achievements.

Comparing Tables 3a and b, the MSC spectral processing has poor efficiency in improving the regression model compared to that based on the original spectra, which may be because the original immersion-based spectra showed higher reproducibility. In the case of the second derivative spectra, the best model performance was obtained in the region 7300–5500  $\text{cm}^{-1}$ , as shown in Figure 4c. Similar to the MSC treatment, most of the region produced

unaltered or worse calibration models than the original spectra (Table 3). Thus, these investigations demonstrate that it is unnecessary to use spectral treatments to construct a better regression model in this system.

The calibration models obtained from the liquid analysis exhibited lower performance compared to those from the solid analysis. The potential of enzymatic hydrolysis was determined to be related to the quantities of individual chemical components and microstructures that form the polymer networks, macromolecule linkages, and cellulose crystalline structures in the pretreated biomass. The reason for the better performance of the solid analysis method compared to the liquid can be due to the chemical and structural information contained in the residue, whereas the NIR spectra from the liquid samples can explain only the quantitative chemical information of the residues by means of subtraction through use of the filtrate. Moreover, in comparison with solid analysis (Figure 2), calibration models using liquid samples showed the tendency of the loss of linear relationships between NIR spectroscopy and wet chemistry (Figure 4). It is unclear whether this phenomenon was attributed to the detection limit of immersion probe or other factors, and therefore research along this line is in progress and will be reported elsewhere.

#### 4. Conclusion

NIR spectroscopy combined with multivariate analysis successfully predicted biomass properties for ethanol production using the residues as well as filtrates obtained after hydrothermal pretreatment of *E. globulus*. When stable and reproducible original spectra are collected, it is not necessary to employ spectral pretreatments to obtain a reliable calibration model. The methodology employing conventional solid samples is quite predictable, but requires sample preparation that translates to “at-line” monitoring in terms of the manufacturing process. On the other hand, liquid analysis can skip these time-consuming steps, and generate a calibration model with reliable predictive ability by using an immersion

probe. This analysis enables the immediate detection of unsatisfactory pretreatments or undesirable contamination, such as that by bacteria, viruses, and certain chemicals, and corresponds to “in-line” monitoring within the manufacturing control system to support stable ethanol conversion.

### **Acknowledgments**

This study was funded by the New Energy and Industrial Technology Development Organization (NEDO).

## References

- [1] L.R. Schimleck, R. Evans, P.D. Jones, R.F. Daniels, G.F. Peter, A. Clark, Estimation of microfibril angle and stiffness by near infrared spectroscopy using sample sets having limited wood density variation, *IAWA J.* 26 (2005) 175–187.
- [2] T. Inagaki, Y. Shinoda, M. Miyazawa, H. Takamura, S. Tsuchikawa, Near-infrared spectroscopic assessment of contamination level of sewage, *Water Sci. Technol.* 61 (2010) 1957–1963.
- [3] M.A. Sanderson, F. Agblevor, M. Collins, D.K. Johnson, Compositional analysis of biomass feedstocks by near infrared reflectance spectroscopy, *Biomass Bioenrg.* 11 (1996) 365–370.
- [4] E.J. Wolfrum, A.D. Sluiter, Improved multivariate calibration models for corn stover feedstock and dilute-acid pretreated corn stover, *Cellulose* 16 (2009) 567–576.
- [5] M. Monrroy, J. R. Garcia, E. Troncosoc, J. Freer, Fourier transformed near infrared (FT-NIR) spectroscopy for the estimation of parameters in pretreated lignocellulosic materials for bioethanol production, *J. Chem. Technol. Biotechnol.* In press (2014) DOI 10.1002/jctb.4427.
- [6] L. Liu, X.P. Ye, A.R. Womac, S. Sokhansanj, Variability of biomass chemical composition and rapid analysis using FT-NIR techniques, *Carbohydr. Polym.* 81 (2010) 820–829.
- [7] X.P. Ye, L. Liu, D. Hayes, A. Womac, K.L. Hong, S. Sokhansanj, Fast classification and compositional analysis of corn stover fractions using Fourier transform near-infrared techniques, *Bioresour. Technol.* 99 (2008) 7323–7332.
- [8] C. Krongtaew, K. Messner, T. Ters, K. Fackler, Characterization of key parameters for biotechnological lignocellulose conversion assessed by FT-NIR spectroscopy. Part II: Quantitative analysis by partial least squares regression, *Bioresources* 5 (2010) 2081–2096.
- [9] Y. Horikawa, T. Imai, R. Takada, T. Watanabe, K. Takabe, Y. Kobayashi, J. Sugiyama, near-infrared chemometric approach to exhaustive analysis of rice straw pretreated for bioethanol conversion, *Appl. Biochem. Biotechnol.* 164 (2011) 194–203.
- [10] Y. Horikawa, T. Imai, R. Takada, T. Watanabe, K. Takabe, Y. Kobayashi, J. Sugiyama, Chemometric analysis with near-infrared spectroscopy for chemically pretreated *Erianthus* toward efficient bioethanol production, *Appl. Biochem. Biotechnol.* 166 (2012) 711–721.
- [11] J.W. Hall, B. McNeil, M.J. Rollins, I. Draper, B.G. Thompson, G. Macaloney, Near-infrared spectroscopic determination of acetate, ammonium, biomass, and glycerol in an industrial *Escherichia coli* fermentation, *Appl. Spectrosc.* 50 (1996) 102–108.
- [12] L.E. Rodriguez-Saona, F.S. Fry, M.A. McLaughlin, E.M. Calvey, Rapid analysis of sugars in fruit juices by FT-NIR spectroscopy, *Carbohydr. Res.* 336 (2001) 63–74.
- [13] D. Cozzolino, M.J. Kwiatkowski, M. Parker, W.U. Cynkar, R.G. Dambergs, M. Gishen, M.J. Herderich, Prediction of phenolic compounds in red wine fermentations by visible and near infrared spectroscopy, *Anal. Chim. Acta.* 513 (2004) 73–80.

- [14] B. Liebmann, A. Friedl, K. Varmuza, Determination of glucose and ethanol in bioethanol production by near infrared spectroscopy and chemometrics, *Anal. Chim. Acta.* 642 (2009) 171–178.
- [15] B. Liebmann, A. Friedl, K. Varmuza, Applicability of near-infrared spectroscopy for process monitoring in bioethanol production, *Biochem. Eng. J.* 52 (2010) 187–193.
- [16] B. Adney, J. Baker, Measurement of cellulase activities, Technical Report NREL/ TP-510-42628 (1996).
- [17] W.W. Luchsinger, R.A. Cornesky, Reducing power by dinitrosalicylic acid method, *Anal. Biochem.* 4 (1962) 346–347.
- [18] A. Savitzky, M.J.E. Golay, Smoothing + differentiation of data by simplified least squares procedures, *Anal. Chem.* 36 (1964) 1627–1639.
- [19] P.C. Williams, D.C. Sobering, Comparison of commercial near-infrared transmittance and reflectance instruments for analysis of whole grains and seeds, *J. Near-Infrared Spec.* 1 (1993) 25–32.
- [20] S. Tsuchikawa, H.W. Siesler, Near-infrared spectroscopic monitoring of the diffusion process of deuterium-labeled molecules in wood. Part II: Hardwood, *Appl. Spectrosc.* 57 (2003) 675–681.
- [21] S. Tsuchikawa, H. W. Siesler, Near-infrared spectroscopic monitoring of the diffusion process of deuterium-labeled molecules in wood. Part I: Softwood. *Appl. Spectrosc.* 57 (2003) 667–674.
- [22] S. Tsuchikawa, H. Yonenobu, H.W. Siesler, Near-infrared spectroscopic observation of the ageing process in archaeological wood using a deuterium exchange method, *Analyst* 130 (2005) 379–384.
- [23] K. Mitsui, T. Inagaki, S. Tsuchikawa, Monitoring of hydroxyl groups in wood during heat treatment using NIR spectroscopy, *Biomacromolecules* 9 (2008) 286–288.
- [24] A. W. Canaza-Cayo, D. Cozzolino, D. Alomar, E. Quispe, A feasibility study of the classification of Alpaca (*Lama pacos*) wool samples from different ages, sex and color by means of visible and near infrared reflectance spectroscopy, *Comput. Electron. Agr.* 88 (2012) 141–147.
- [25] T. Kawai, H. Nakazawa, N. Ida, H. Okada, S. Tani, J. Sumitani, T. Kawaguchi, W. Ogasawara, Y. Morikawa, Y. Kobayashi, Analysis of the saccharification capability of high-functional cellulase JN11 for various pretreated biomasses through a comparison with commercially available counterparts, *J. Ind. Microbiol. Biotechnol.* 39 (2012) 1741–1749.
- [26] W.S.L. Mok, M.J. Antal, Uncatalyzed solvolysis of whole biomass hemicellulose by hot compressed liquid water, *Ind. Eng. Chem. Res.* 31 (1992) 1157–1161.
- [27] N. Mosier, C. Wyman, B. Dale, R. Elander, Y.Y. Lee, M. Holtzapple, M. Ladisch, Features of promising technologies for pretreatment of lignocellulosic biomass, *Bioresour. Technol.* 96 (2005) 673–686.



Table 1. Statistical summary of the saccharification data for the calibration and prediction sets.

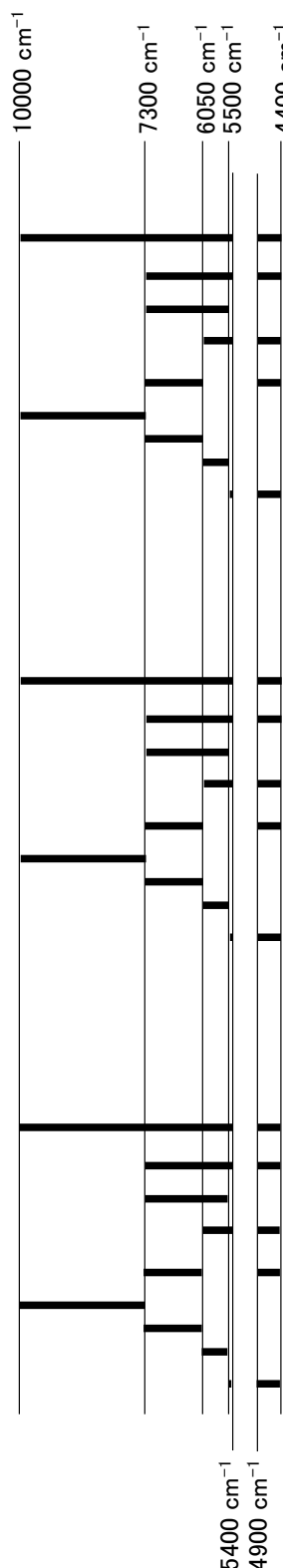
	Calibration set					Prediction set				
	Samples	Max. (%)	Min. (%)	Mean (%)	S. D. (%)	Samples	Max. (%)	Min. (%)	Mean (%)	S. D. (%)
Saccharification ratio (%)	54	62.36	4.95	27.48	16.82	18	61.49	5.80	29.42	17.76



Table 2. Statistical summary of the calibration model obtained from solid disk sample and applied to the prediction set. The models were obtained by employing the original (a), MSC-treated (b), and second-derivative (c) spectra. A schematic illustration has been included on the left for assistance in understanding each spectral region.

10000 cm <sup>-1</sup>	7300 cm <sup>-1</sup>	6050 cm <sup>-1</sup>	5500 cm <sup>-1</sup>	4000 cm <sup>-1</sup>	<b>a</b>						
					PLS factors	Calibration set		Prediction set			
						R <sup>2</sup> <sub>c</sub>	RMSEC(%)	R <sup>2</sup> <sub>p</sub>	RMSEP(%)	RPD	
					10000 – 4000 cm <sup>-1</sup>	4	0.97	3.07	0.99	2.08	8.55
					7300 – 4000 cm <sup>-1</sup>	4	0.94	4.04	0.93	4.52	3.93
					7300 – 5500 cm <sup>-1</sup>	3	0.96	3.42	0.98	2.57	6.91
					6050 – 4000 cm <sup>-1</sup>	4	0.96	3.53	0.97	3.10	5.73
					7300 – 6050						
					5500 – 4000 cm <sup>-1</sup>	4	0.94	4.18	0.93	4.64	3.83
					10000 – 7300 cm <sup>-1</sup>	4	0.97	2.93	0.98	2.65	6.69
					7300 – 6050 cm <sup>-1</sup>	3	0.92	4.71	0.97	3.28	5.42
					6050 – 5500 cm <sup>-1</sup>	4	0.97	2.77	0.97	2.80	6.34
					5500 – 4000 cm <sup>-1</sup>	4	0.96	3.52	0.97	3.23	5.50
					<b>b</b>						
					PLS factors	Calibration set		Prediction set			
						R <sup>2</sup> <sub>c</sub>	RMSEC(%)	R <sup>2</sup> <sub>p</sub>	RMSEP(%)	RPD	
					10000 – 4000 cm <sup>-1</sup>	2	0.97	2.96	0.99	2.08	8.54
					7300 – 4000 cm <sup>-1</sup>	2	0.96	3.25	0.98	2.57	6.92
					7300 – 5500 cm <sup>-1</sup>	2	0.97	2.85	0.99	2.15	8.26
					6050 – 4000 cm <sup>-1</sup>	3	0.97	3.08	0.97	2.84	6.26
					7300 – 6050						
					5500 – 4000 cm <sup>-1</sup>	2	0.97	2.94	0.98	2.31	7.70
					10000 – 7300 cm <sup>-1</sup>	2	0.97	3.11	0.98	2.59	6.84
					7300 – 6050 cm <sup>-1</sup>	2	0.97	2.89	0.99	2.14	8.30
					6050 – 5500 cm <sup>-1</sup>	3	0.97	3.04	0.95	4.10	4.34
					5500 – 4000 cm <sup>-1</sup>	3	0.96	3.21	0.97	3.23	5.49
					<b>c</b>						
					PLS factors	Calibration set		Prediction set			
						R <sup>2</sup> <sub>c</sub>	RMSEC(%)	R <sup>2</sup> <sub>p</sub>	RMSEP(%)	RPD	
					10000 – 4000 cm <sup>-1</sup>	2	0.95	3.67	0.96	3.45	5.15
					7300 – 4000 cm <sup>-1</sup>	2	0.95	3.69	0.96	3.46	5.14
					7300 – 5500 cm <sup>-1</sup>	2	0.97	3.07	0.98	2.62	6.77
					6050 – 4000 cm <sup>-1</sup>	2	0.95	3.87	0.96	3.67	4.84
					7300 – 6050						
					5500 – 4000 cm <sup>-1</sup>	2	0.95	3.79	0.96	3.58	4.96
					10000 – 7300 cm <sup>-1</sup>	3	0.96	3.37	0.87	6.26	2.84
					7300 – 6050 cm <sup>-1</sup>	3	0.98	2.50	0.96	3.42	5.20
					6050 – 5500 cm <sup>-1</sup>	2	0.97	2.86	0.98	2.23	7.96
					5500 – 4000 cm <sup>-1</sup>	2	0.94	3.99	0.95	3.82	4.65

Table 3 Statistics summary of the calibration model obtained from filtrate measured by immersion probe and applied to the prediction set. The models were obtained by employing the original (a), MSC-treated (b), and second-derivative (c) spectra. A schematic illustration has been included on the left for assistance in understanding each spectral region.



**a**

	PLS factors	Calibration set		Prediction set		
		$R^2_c$	RMSEC(%)	$R^2_p$	RMSEP(%)	RPD
10000 – 5400						
4900 – 4400 $\text{cm}^{-1}$	7	0.95	3.97	0.94	4.45	3.99
7300 – 5400						
4900 – 4400 $\text{cm}^{-1}$	6	0.95	3.81	0.93	4.55	3.90
7300 – 5500 $\text{cm}^{-1}$	7	0.95	3.86	0.86	6.47	2.74
6050 – 5400						
4900 – 4400 $\text{cm}^{-1}$	4	0.86	6.31	0.86	6.60	2.69
7300 – 6050						
4900 – 4400 $\text{cm}^{-1}$	4	0.89	5.53	0.87	6.38	2.78
10000 – 7300 $\text{cm}^{-1}$	4	0.41	13.06	0.27	15.07	1.18
7300 – 6050 $\text{cm}^{-1}$	7	0.96	3.42	0.89	5.81	3.06
6050 – 5500 $\text{cm}^{-1}$	5	0.63	10.34	0.68	10.00	1.78
5500 – 5400						
4900 – 4400 $\text{cm}^{-1}$	6	0.89	5.71	0.90	5.63	3.16

**b**

	PLS factors	Calibration set		Prediction set		
		$R^2_c$	RMSEC(%)	$R^2_p$	RMSEP(%)	RPD
10000 – 5400						
4900 – 4400 $\text{cm}^{-1}$	6	0.94	4.14	0.94	4.25	4.18
7300 – 5400						
4900 – 4400 $\text{cm}^{-1}$	6	0.94	4.07	0.92	5.03	3.53
7300 – 5500 $\text{cm}^{-1}$	3	0.88	5.94	0.90	5.42	3.28
6050 – 5400						
4900 – 4400 $\text{cm}^{-1}$	3	0.86	6.46	0.90	5.70	3.12
7300 – 6050						
4900 – 4400 $\text{cm}^{-1}$	7	0.95	3.73	0.93	4.58	3.88
10000 – 7300 $\text{cm}^{-1}$	3	0.82	7.15	0.66	10.22	1.74
7300 – 6050 $\text{cm}^{-1}$	4	0.89	5.73	0.91	5.31	3.34
6050 – 5500 $\text{cm}^{-1}$	3	0.49	12.15	0.08	16.88	1.05
5500 – 5400						
4900 – 4400 $\text{cm}^{-1}$	6	0.89	5.62	0.86	6.55	2.71

**c**

	PLS factors	Calibration set		Prediction set		
		$R^2_c$	RMSEC(%)	$R^2_p$	RMSEP(%)	RPD
10000 – 5400						
4900 – 4400 $\text{cm}^{-1}$	6	0.91	4.96	0.88	6.00	2.96
7300 – 5400						
4900 – 4400 $\text{cm}^{-1}$	6	0.91	5.22	0.88	5.98	2.97
7300 – 5500 $\text{cm}^{-1}$	5	0.93	4.49	0.89	5.83	3.05
6050 – 5400						
4900 – 4400 $\text{cm}^{-1}$	2	0.74	8.60	0.80	7.81	2.27
7300 – 6050						
4900 – 4400 $\text{cm}^{-1}$	3	0.81	7.47	0.85	6.80	2.61
10000 – 7300 $\text{cm}^{-1}$	6	0.68	9.60	0.46	12.96	1.37
7300 – 6050 $\text{cm}^{-1}$	3	0.86	6.26	0.86	6.60	2.69
6050 – 5500 $\text{cm}^{-1}$	5	0.60	10.78	0.60	11.05	1.61
5500 – 5400						
4900 – 4400 $\text{cm}^{-1}$	3	0.76	8.35	0.80	7.88	2.25

## Figure legends

Fig. 1

NIR spectra obtained from solid samples by sphere diffuse reflectance (a–c) and from liquid samples by immersion probe (d–f) are shown.

(a) and (d) are original, (b) and (e) are MSC-treated, and (c) and (f) are second-derivative spectra. The nine spectra were obtained from hydrothermal treatment at 190°C for various times (1–64 min).

Fig. 2

Relationships between the saccharification ratio measured by DNS method and predicted by NIR spectroscopy obtained from solid disk samples. Open circles are estimated from calibration set and filled those are from prediction set. The correlations were obtained from (a) original, (b) MSC-treated, and (c) second-derivative spectra.

Fig. 3

Regression coefficient for the calibration model of the solid analysis in the range 10000–4000  $\text{cm}^{-1}$ . The lines at 7108 and 5224  $\text{cm}^{-1}$  were assigned to absorbed water; at 6900 and 5980  $\text{cm}^{-1}$  to lignin; and at 5810  $\text{cm}^{-1}$  to hemicelluloses. Regression coefficients were computed from (a) original, (b) MSC-treated, and (c) second-derivative spectra.

Fig. 4

Relationships between the saccharification ratio measured by DNS method and predicted by NIR spectroscopy obtained from the filtrates. Open circles are estimated from calibration set and filled those are from prediction set. The correlations were obtained from (a) original, (b) MSC-treated, and (c) second-derivative spectra.

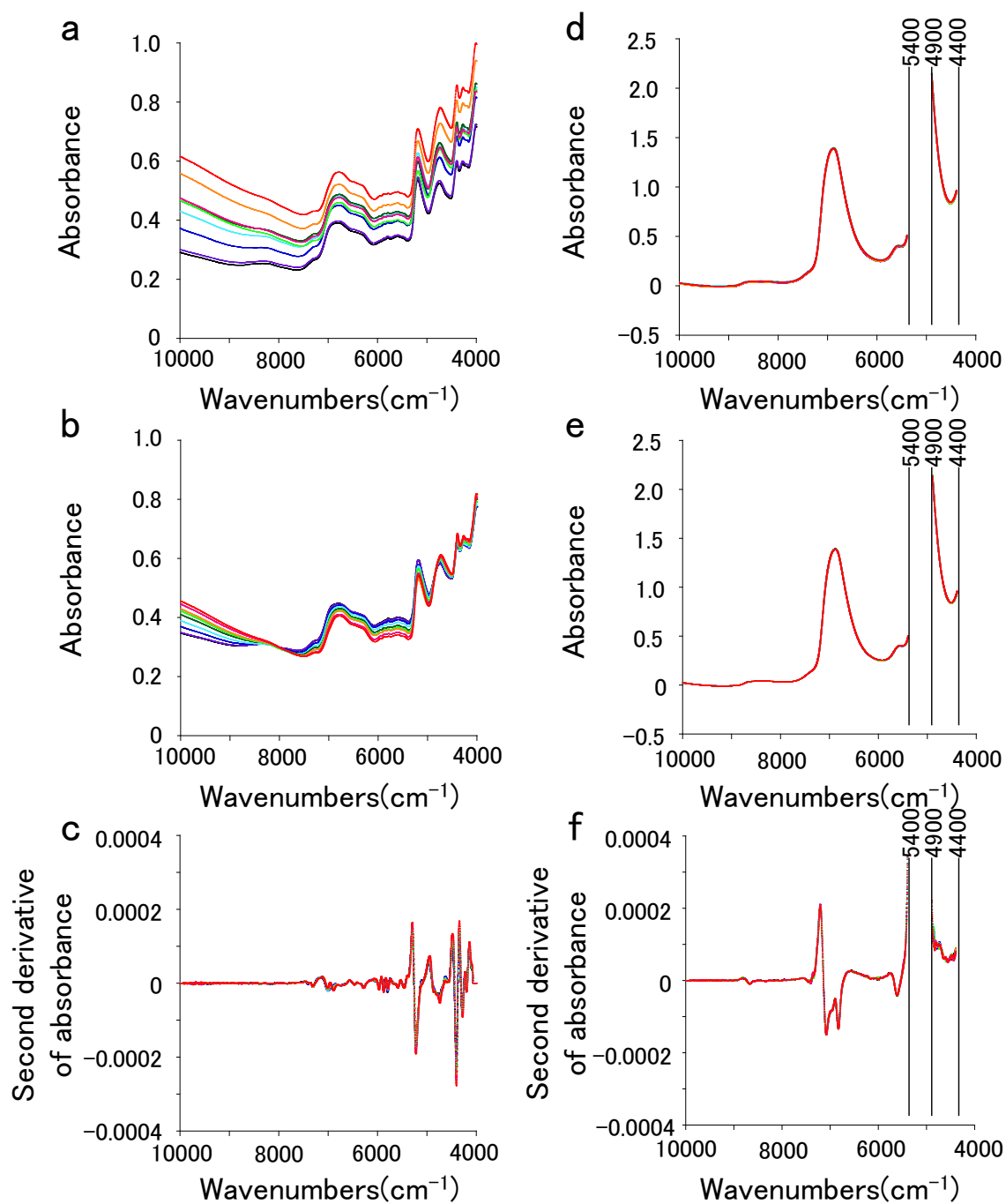


Fig. 1

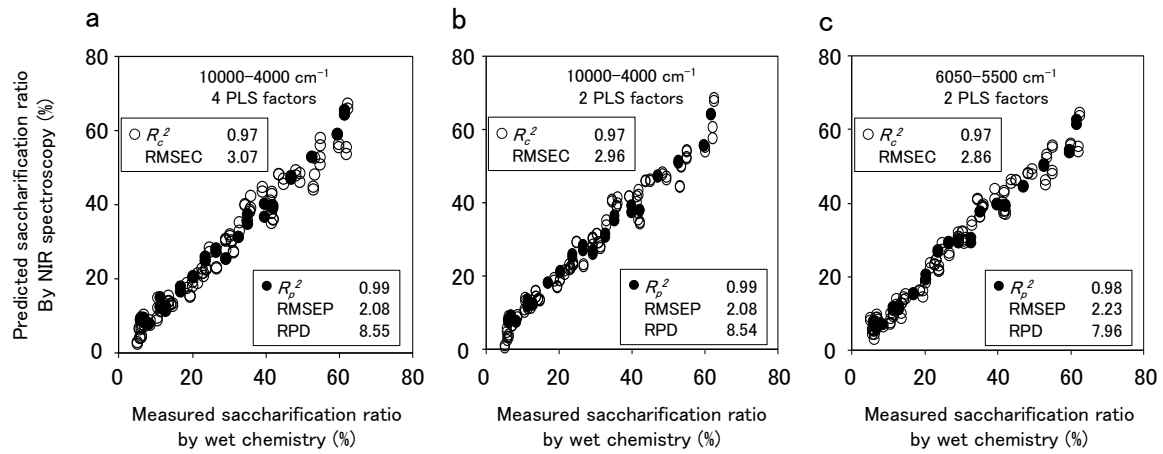


Fig. 2

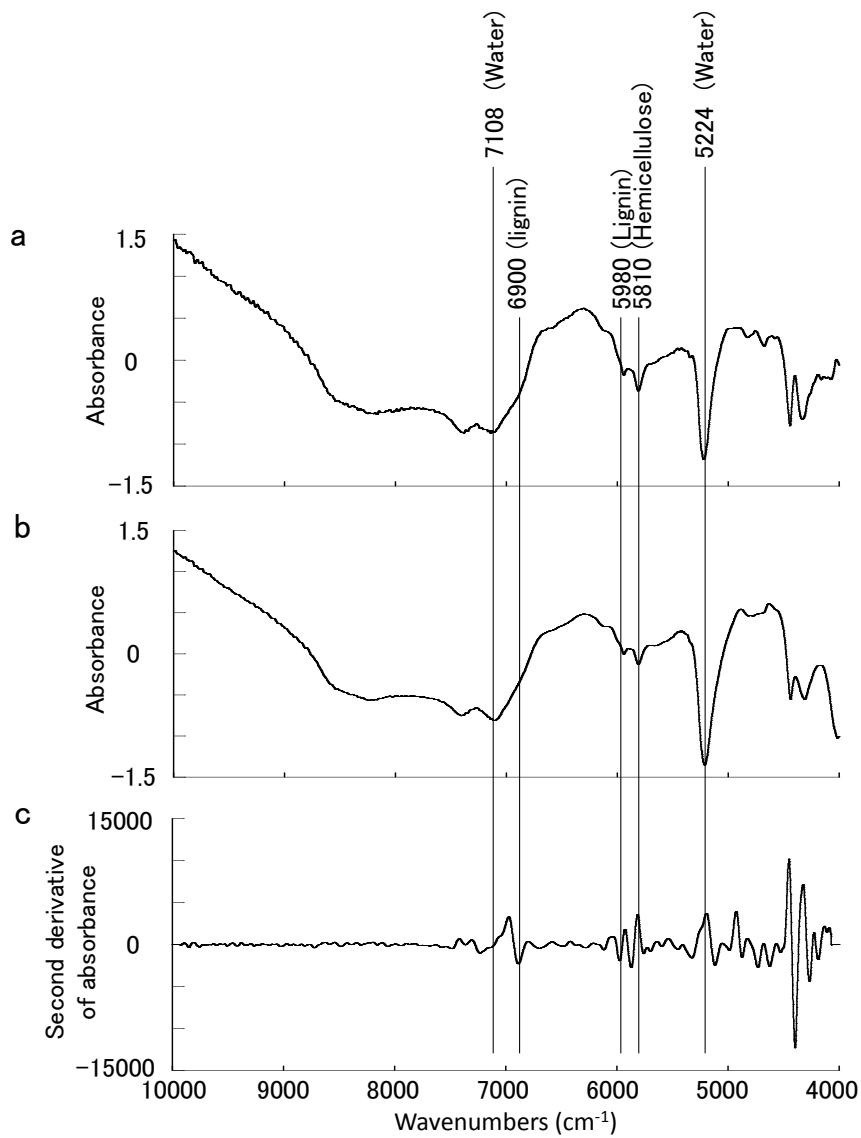


Fig. 3

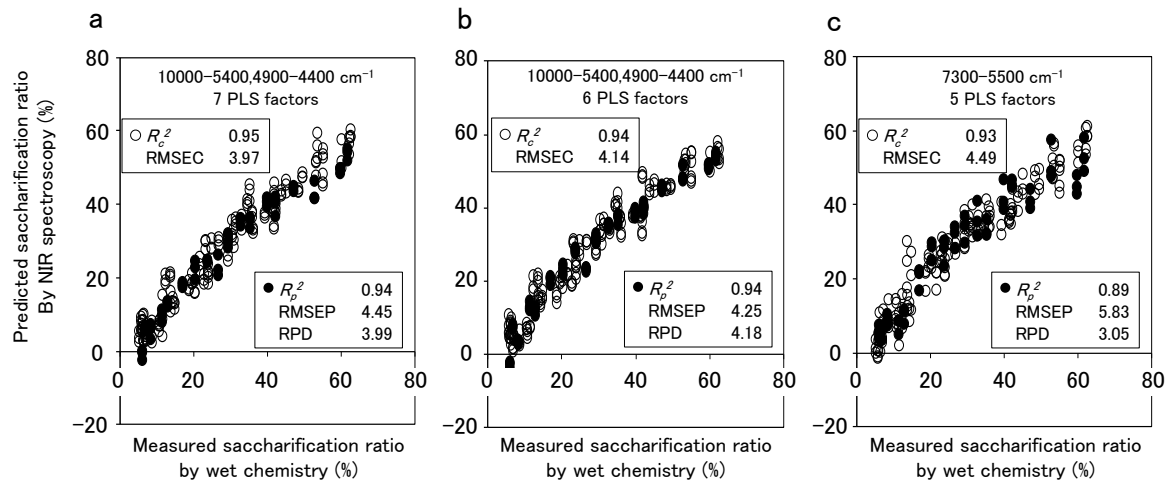


Figure 4